Drivers' Manoeuvre Classification for Safe HRI

Erwin Jose Lopez Pulgarin[1](0000-0001-9927-6688), Guido Herrmann[2](0000-0001-5390-4538), and Ute Leonards[3](0000-0001-6143-7466)

[1] Mechanical Engineering, University of Bristol, Bristol, UK
erwin.lopez@bristol.ac.uk,
[2] Mechanical Engineering, University of Bristol, Bristol, UK
g.herrmann@bristol.ac.uk,
[3] Experimental Psychology, University of Bristol, Bristol, UK
ute.leonards@bristol.ac.uk

Abstract. Ever increasing autonomy of machines and the need to interact with them creates challenges to ensure safe operation. Recent technical and commercial interest in increasing autonomy of vehicles has led to the integration of more sensors and actuators inside the vehicle, making them more like robots. For interaction with semi-autonomous cars, the use of these sensors could help to create new safety mechanisms. This work explores the concept of using motion tracking (i.e skeletal tracking) data gathered from the driver whilst driving to learn to classify the manoeuvre being performed. A kernel-based classifier is trained with empirically selected features based on data gathered from a Kinect V2 sensor in a controlled environment. This method shows that skeletal tracking data can be used in a driving scenario to classify manoeuvres and sets a background for further work.

Keywords: HRI, semi-autonomous vehicles, vehicles, driver actions, classification, machine learning

1 Introduction

Recent trends in automotive driver assist systems point towards cars becoming more robot-like. Advanced sensing and actuating capabilities allow for increased autonomy in the form of advanced driver assistance systems (ADAS) (e.g. Lane Keeping Assistant, Cruise Control, etc) and permit some basic autonomous navigation, while recent commercial efforts have pushed for fully autonomous operation of passenger vehicles [4].

An increasing trend when having an automated method dealing or interacting with users is to use physiological measurements (i.e. signals measured from a person related to mind and body state) in order to get insight into the user's inner states like stress levels [5], which could prove to be useful in different applications like semi-automatic driving [7]. Among these measurements, movement information or skeletal tracking data has become very popular due to low-cost sensors that enable to track human position and that permit its use in different indoor scenarios [8] [17].

Recent advances in ADAS systems [13] and estimation techniques using advanced sensor input from vehicles [8] [18] [2] show the multiple possibilities of implementing vision-based solutions for in-cabin operation to observe a driver's action like hand posture recognition[18], in-car movement [8] and general human pose estimation[2]. However, the use of skeletal tracking data has been fairly limited.

In this paper, a classification method for driver manoeuvres is developed using a data-driven approach with a Support Vector Machine (SVM) classifier, including skeletal tracking and driver input data. First, the main techniques are explained in Sections 2. Section 3 explains the results and Section 4 talks about conclusions and future work.

2 Skeletal and in-car sensor information

Limb position and movement are related to the final output of a decision making process, and provide information about complex, long repetitive movements or motion patterns that could be quickly identified.

Studies [10] have shown that, for a trained and controlled task such as driving, the muscles, i.e. the general movement of the arms, do not differ between test subjects with different driving experiences. This is a strong indication about movement repeatability whilst driving and it is thus possible to use body movement information as a measured quantity.

Sensors that acquire colour (RGB) data from the environment are usually called RGB cameras, and have enabled numerous developments by exploiting the 2D representation of the amount of light reflected in a 3D scene; RGB-D cameras provide information about the distance between sensor and the scene being recorded, which allows for the development of faster, more precise classification and recognition methods like human body detection and tracking [6] [1] [16] [14].

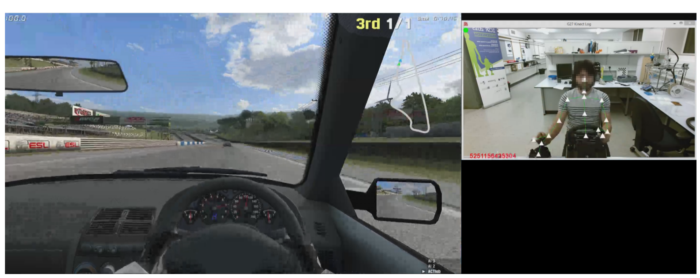

Fig. 1: Experimental setup for data acquisition whilst driving. driving simulator output(left), recording screen with skeletal tracking(right))

Data acquired from a set of driving experiments is presented. Thus, a controlled simulated environment was selected, providing a trade-off between repeatability and ease in implementation. A driving simulation environment for a UK based car (left lane driving) is implemented, using a gaming-grade user input device (Logitech G27), a semi-professional racing simulator (Live for Speed) and a Kinect V2 sensor (see Figure 1). The use of the Kinect V2 for body movement observation creates technical challenges whilst working in a car-like environment [8] [17]. In our case, the Kinect V2 faces down, pointing towards the driver to achieve a better view (see Figure 2), with a 70 cm distance in line of sight from the sensor to the steering wheel. The overall arrangement attempts to recreate a similar arrangement as in a road vehicle (although inner spacing is not ideal, as the shift stick is at the same level as the steering wheel's and the steering wheel vertical position is 10 cm higher than in normal vehicles).

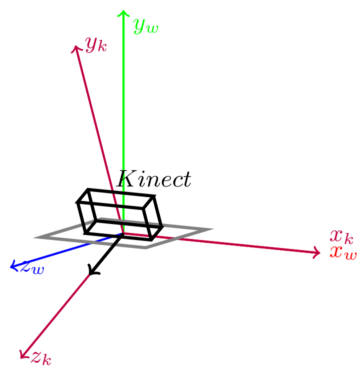

Fig. 2: Kinect Coordinate frame (x_k, y_k, z_k) related to World Coordinate frame (x_w, y_w, z_w) (rotated in x axes)

Data was acquired from six test subjects during two driving experiments using different driving styles (i.e. drive as fast as possible and driving carefully on the track) with 5 turns per side (i.e. left turn, right turn) during 2 minutes each in a racing track and a data sampling of 30 ms. Two different driving styles were used to roughly take into account the difference in arms movement whilst driving at different speeds. Duration of turns can be from 0.5 to 2 seconds, getting between 150 and 600 data points per turn. Hence, these data allow to model their behaviour during different manoeuvres.

After testing the geometrical and operational constraint of the sensor and the experimental setup, some empirically selected features based on limb angle position, combined with the driver's input, have been used as defined below:

- **azimuthEL:** Azimuth angle of the spherical projection of the angle between the left elbow and the left shoulder.
- **elevationEL:** Elevation angle of the spherical projection of the angle between the left elbow and the left shoulder.

- **azimuthER:** Azimuth angle of the spherical projection of the angle between the right elbow and the right shoulder.
- **elevationER:** Elevation angle of the spherical projection of the angle between the right elbow and the right shoulder.
- **BackLean:** Difference between torso and back position in z axes.
- **dCenterX:** Difference between left hand position and torso in x axes.
- **Steering Wheel Angle:** Turning angle recorded by the Logitech G27 device.

These features were selected from a set of 95 signals. They were a trade-off between numerical load of the machine learning approach and success rate in feature detection.

2.1 Data-driven techniques

Data-driven techniques have proven useful to classify data for vehicle manoeuvres and other related scenarios [9].

In the area of common classifiers used in data-driven techniques (e.g. SVM, neural networks, random trees), kernel-based methods such as Support Vector Machines (SVM) provide a high flexibility of generalizing linear and non-linear processes whilst maintaining small training and execution periods. Hence, SVM is used in this case. SVM theory was initially developed by V. Vapnik in the early 1980s and focused on binary classification problems, using the paradigm of statistical learning theory. The basic idea behind SVM is to find an optimal hyperplane that maximizes the separation margin between classes. Finding this hyperplane is equivalent to solving a constrained optimization problem whose solution is a linear combination of training examples that are located outside the region that lies between what is known as the support vectors [3].

In our specific scenario, a state-transition model (i.e Markov chain) of the driving process is set up with a transition probability equal for all states. Three states are proposed, with "straight" being the initial state and "left turn" or "right turn" the possible next states, always having to go back to the initial state to transition between turning. The reduced set of states and the equal probability for transitions restriction simplify a possibly more complex state model, avoids coupling between manoeuvres and establishes an intuitive way of performing basic turning movements, that can lead to more complex manoeuvres (see Figure 3).

Based on a data-driven approach to model the problem, and using the state transition explained in Figure 3, a classification scheme using three classes is designed. The scheme estimates the current manoeuvre performed between 3 possible types of manoeuvre (turn left, turn right, straight movement), with a feature vector only using data from one time-step (i.e. data at time t) (see Figure 4).

Labels corresponding to the state transition model in Figure 4 (i.e. left turn, right turn and straight) are given manually to segments of the recorded data,

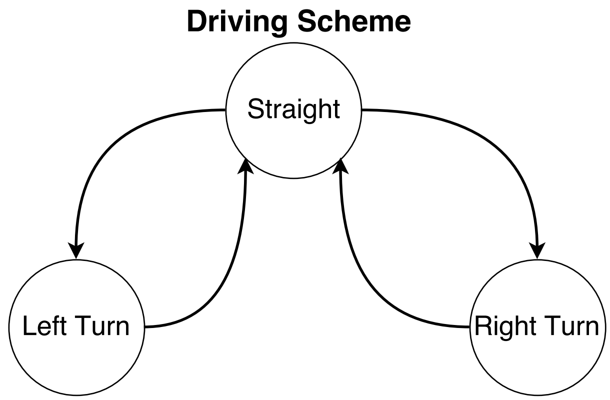

Fig. 3: Driving manoeuvres modelling with basic transitions.

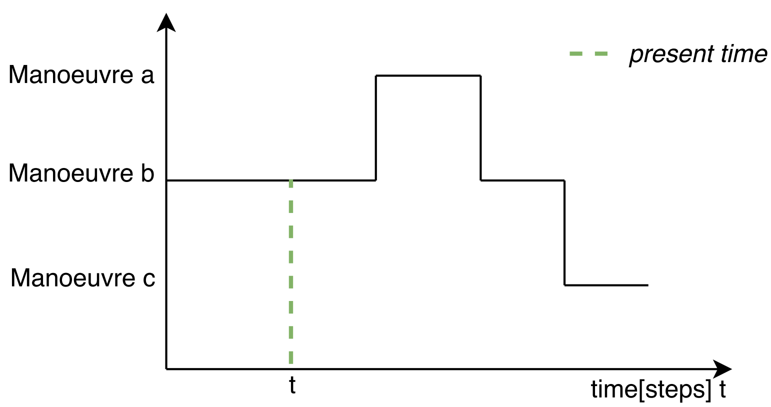

Fig. 4: Driving manoeuvres classification description in time.

based on the position on the track and the steering wheel angle seen on the video recordings of the driver.

Classification results were evaluated with respect to metrics for binary classifiers, namely $Precision$, $Recall$ and $F1$ score. These are based on the number of true-positives tp (i.e. predicted true, expected true), true-negatives tn (i.e. predicted false, expected false), false-positives fp (i.e. predicted true, expected false) and false-negatives fn (i.e. predicted false, expected true) (see [15] for a general definition).

$$Precision = \frac{tp}{tp + fp} \qquad Recall = \frac{tp}{tp + fn} \qquad F1 \quad = 2\frac{Precision * Recall}{Precision + Recall} \tag{1}$$

Precision represents the repeatability of the prediction or how many predictions are relevant, recall represents how many relevant predictions are done and F1 is the harmonic mean between precision and recall to represent a balanced predictor.

As a general objective, a high precision (no prediction left undone), high recall (high probability of having a good prediction) and a close to 1 F1 Score is desired.

The model was trained and tested from the above explained experimental data with a stratified (i.e. separated by classes) cross-validated set of 25% testing data, 75% training data. This means that the dataset from the test trials is divided into the 3 available classes. Each class has about 60 intervals of data of different temporal lengths. The data from every class is randomly selected and divided into training and testing data.

The dataset is filtered with different combinations of common skeletal tracking and pre-processing filtering techniques [12]: raw data, data filtered by a Double Exponential Smoothing Filter, the same filtered data with normalization by dimensions mapping the min and max values of a given dimension to 0 and 1 (i.e. hard whitening) and the same previous filtered data with normalization by dimensions subtracting the mean of the values and divide by twice the standard deviation (i.e. soft whitening). The data used in this learning process are those 7 signals explained at the end of the previous section. An SVM classifier with a multiclass strategy of one-vs-one is used with standardized input, radial based kernel function and ISDA solver.

3 Results

Performance metrics averaged from the results of all 6 test subjects can be seen in Table 1, with "NoFilter" showing the results with non-filtered training data, "Filter" showing the results with double-exponential filter training data, "FilterHW" showing the results with filtered + hard-whitened training data and "FilterSW" showing the results with filtered + soft-whitened training data. The used feature vector, together with the selected classification algorithm, allow us to discriminate between various manoeuvres whilst generalizing throughout different drivers, successfully classify it.

Mean performance metrics are all above 85% without filtered training data and above 90% with filtered data, including the F1 metric which is above 90% in all cases; the F1 metric shows a balanced performance between missed classifications and true classifications, as can be seen in Figure 5, with low numbers of missed classifications throughout the tests whilst remaining sensible to changes.

The proposed method is able to learn to classify the 3 manoeuvres using a relatively small dataset per test subject, exploiting the repeatability of arm movement before and whilst performing a driving manoeuvre, being the main advantage compared to other methods that require big training sets to obtain performance over 85%.

Table 1: Manoeuvre Classification performance metrics

		NoFilter	Filter	FilterHW	FilterSW
Turn Left	Precision	0.8617	0.9259	0.9247	0.7111
	Recall	0.9709	0.9841	0.9849	0.9418
	F1	0.9130	0.9541	0.9538	0.8104
Straight	Precision	0.9878	0.9901	0.9946	0.9764
	Recall	0.9266	0.9617	0.9570	0.8711
	F1	0.9562	0.9757	0.9754	0.9208
Turn Right	Precision	0.9461	0.9747	0.9677	0.9141
	Recall	0.9663	0.9794	0.9828	0.9314
	F1	0.9561	0.9770	0.9752	0.9227

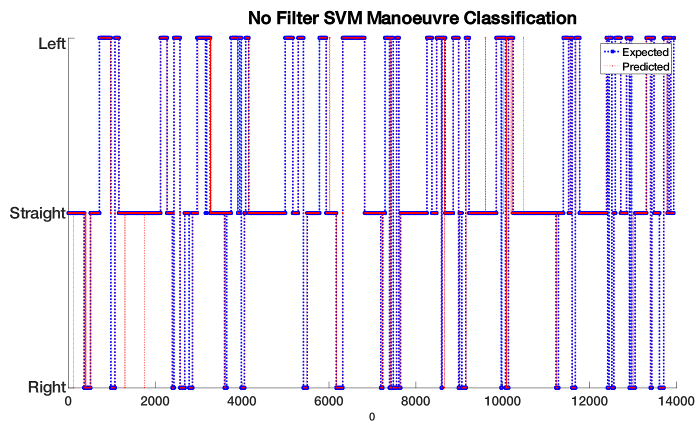

Fig. 5: Manoeuvres classification for all test subjects.)

4 Conclusions

Current implementation shows that empirically selected features based on skeletal tracking information and driver input are sufficient to create models that classify the type of manoeuvre being performed by a driver, using both filtered and unfiltered training data. This scheme shows the ability of skeletal-tracking-based features to generalize a movement, being able to classify driving manoeuvres with relatively small datasets, compared to the usual big datasets required for classification tasks.

The generated scheme was general enough to classify new data from the test subjects, but not enough to classify data from people whose driving style is unknown (i.e model sensitive to training data). More information could prove to be useful into creating richer models but consideration must be taken due to limitation of kernel-based approaches like the one used (e.g. the hyperplane or high dimension description could grow too big or too complex that it's infeasible to separate).

Future work will focus on using more sensors to acquire driver-related information and enrich our understanding of the driver's inner model. We will also look into creating richer driving scenarios that allow to simulate different mental workloads or distraction levels, in order to know how driver behaviour changes during manoeuvres when affected by different levels of distractions.

Data relevant to the research results is openly available at the time of publication [11].

References

1. Bo, L., Ren, X., Fox, D.: Unsupervised Feature Learning for RGB-D Based Object Recognition. In: Desai, J.P., Dudek, G., Khatib, O., Kumar, V. (eds.) Experimental Robotics, pp. 387–402. No. 88 in Springer Tracts in Advanced Robotics, Springer International Publishing (2013), http://link.springer.com/chapter/10.1007/978-3-319-00065-7_27, dOI: 10.1007/978-3-319-00065-7_27
2. Crabbe, B., Paiement, A., Hannuna, S., Mirmehdi, M.: Skeleton-Free Body Pose Estimation From Depth Images for Movement Analysis. In: The IEEE International Conference on Computer Vision (ICCV) Workshops (Dec 2015)
3. Duda, R.O., Hart, P.E., others: Pattern classification and scene analysis, vol. 3. Wiley New York (1973)
4. Franke, U., Pfeiffer, D., Rabe, C., Knoeppel, C., Enzweiler, M., Stein, F., Herrtwich, R.: Making Bertha See. In: 2013 IEEE International Conference on Computer Vision Workshops (ICCVW). pp. 214–221 (2013)
5. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems 6(2), 156–166 (Jun 2005)
6. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: RGB-D mapping: Using Kinect-style depth cameras for dense 3d modeling of indoor environments. The International Journal of Robotics Research 31(5), 647–663 (Apr 2012), http://dx.doi.org/10.1177/0278364911434148
7. Koenig, A., Caruso, A., Bolliger, M., Somaini, L., Omlin, X., Morari, M., Riener, R.: Model-based heart rate control during robot-assisted gait training. In: 2011 IEEE International Conference on Robotics and Automation (ICRA). pp. 4151–4156 (May 2011)
8. Kondyli, A., Sisiopiku, V., Barmpoutis, A.: A 3d experimental framework for exploring drivers' body activity using infrared depth sensors. In: 2013 International Conference on Connected Vehicles and Expo (ICCVE). pp. 574–579 (Dec 2013)
9. Kumar, P., Perrollaz, M., Lefèvre, S., Laugier, C.: Learning-based approach for online lane change intention prediction. In: 2013 IEEE Intelligent Vehicles Symposium (IV). pp. 797–802 (Jun 2013)
10. Liu, Y., Ji, X., Ryouhei, H., Takahiro, M., Lou, L.: Function of shoulder muscles of driver in vehicle steering maneuver. Science China Technological Sciences 55(12), 3445–3454 (Sep 2012), http://link.springer.com/article/10.1007/s11431-012-5045-9
11. Lopez Pulgarin, E.J., Herrmann, G., Leonards, U.: Dataset for Drivers' Manoeuvre Classification for Safe HRI (Aug 2017), https://doi.org/10.5281/zenodo.556961

12. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York, 2nd edn. (2006)
13. Perrett, T., Mirmehdi, M., Dias, E.: Cost-based Feature Transfer for Vehicle Occupant Classification. arXiv:1512.07080 [cs] (Dec 2015), http://arxiv.org/abs/1512.07080, arXiv: 1512.07080
14. Ren, X., Bo, L., Fox, D.: RGB-(D) scene labeling: Features and algorithms. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2759–2766 (Jun 2012)
15. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information Processing & Management 45(4), 427–437 (Jul 2009), http://www.sciencedirect.com/science/article/pii/S0306457309000259
16. Spinello, L., Arras, K.O.: People detection in RGB-D data. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 3838–3843 (Sep 2011)
17. Tran, C., Trivedi, M.: Towards a vision-based system exploring 3d driver posture dynamics for driver assistance: Issues and possibilities. In: 2010 IEEE Intelligent Vehicles Symposium (IV). pp. 179–184 (Jun 2010)
18. Zhao, C., Zhang, B., HE, J., Lian, J.: Recognition of driving postures by contourlet transform and random forests. IET Intelligent Transport Systems 6(2), 161–168 (Jun 2012)